

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 13:26:44 ; Search time 39 Seconds  
(without alignments)  
286.183 Million cell updates/sec

Title: US-10-659-782A-32  
Perfect score: 616  
Sequence: 1 MPSPGTVCSLLGLMLDL.....PPSSRRSRSHQSPCSPEL 116

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	198	32.1	117	1 A59316	ghrelin precursor
2	159.5	25.9	117	1 B59316	ghrelin precursor
3	75.5	12.3	573	2 JC4335	anti-mullerian hor
4	72.5	11.8	519	2 G84707	probable MYB famil
5	72.5	11.8	666	2 T22943	hypothetical prote
6	72.5	11.8	725	1 E64211	virulence-associat
7	72	11.7	764	2 I48882	thyrotropin recept
8	70	11.4	307	2 T33503	hypothetical prote
9	69.5	11.3	302	2 H96792	unknown protein F1
10	68.5	11.1	449	2 C39926	hypothetical 51.8K
11	68.5	11.1	456	2 T35474	50kD proline rich
12	68.5	11.1	796	2 T32425	hypothetical prote
13	68	11.0	94	2 B75564	hypothetical prote
14	68	11.0	764	2 A35956	thyrotropin recept
15	67.5	11.0	136	2 AG0449	regulator of nucle
16	67.5	11.0	363	2 F91265	sensor protein Bas
17	67.5	11.0	363	2 C86106	sensor protein for
18	67.5	11.0	363	2 JX0285	sensor protein bas
19	67.5	11.0	406	2 T28957	hypothetical prote
20	67.5	11.0	1704	2 A59188	ATP-binding casset
21	67.5	11.0	1704	2 S71363	probable ATP-bind
22	67	10.9	267	2 AB0461	probable carboxyme
23	67	10.9	381	2 S16506	hypothetical prote
24	67	10.9	749	2 A75560	conserved hypothet
25	66.5	10.8	458	2 AC0216	para-aminobenzoate
26	66.5	10.8	540	2 T27400	hypothetical prote
27	66	10.7	944	2 T41711	probable alpha-tre
28	65.5	10.6	1027	2 B64187	conserved hypothet
29	65	10.6	309	2 S77905	lyase - Pseudomona

30	65	10.6	383	2 A56084	interleukin-lbета
31	65	10.6	490	2 S67581	STP4 protein - yea
32	64.5	10.5	187	2 T51876	hypothetical prote
33	64.5	10.5	354	2 G75548	ABC transporter, A
34	64.5	10.5	467	1 S45493	serine proteinase
35	64.5	10.5	594	2 T15202	hypothetical prote
36	64.5	10.5	761	2 T15776	hypothetical prote
37	64.5	10.5	1001	2 T28897	hypothetical prote
38	64.5	10.5	1487	2 S62048	probable membrane
39	64.5	10.5	6805	2 S20901	titin - rabbit (fr
40	64	10.4	413	2 E82312	phage integrase VC
41	64	10.4	555	2 H83043	hypothetical prote
42	64	10.4	637	2 T03842	fission yeast Skb1
43	63.5	10.3	122	2 JH3144	hypothetical prote
44	63.5	10.3	428	2 JH0634	site-specific DNA-
45	63.5	10.3	522	2 S33029	hypothetical prote

ALIGNMENTS

RESULT 1

A59316  
ghrelin precursor - human  
N;Alternate names: preproghrelin  
C:Species: Homo sapiens (man)  
C>Date: 16-Jun-2000 #sequence\_revision 16-Jun-2000 #text\_change 09-Jul-2004  
C:Accession: A59316  
R;Kojima, M.; Hosoda, H.; Date, Y.; Nakazato, M.; Matsuo, H.; Kangawa, K.  
Nature 402, 656-660, 1999  
A;Title: Ghrelin is a growth-hormone-releasing acylated peptide from stomach.  
A;Reference number: A59316; MUID:20067959; PMID:10604470  
A;Accession: A59316  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-117 <KOJ>  
A;Cross-references: UNIPROT:Q9UBU3; GB:AB029434; NID:G6691571; PIDN:BAA89371.1; PID:G66  
A;Experimental source: tissue stomach endocrine cells  
A;Note: submitted to GenBank, June 1999  
C;Comment: Ghrelin secreted by the stomach stimulates the release of somatotropin (grow  
C;Superfamily: motilin  
C;Keywords: hormone; lipoprotein; stomach  
F;1-23/Domain: signal sequence #status predicted <SIG>  
F;24-51/Product: ghrelin #status predicted <MAT>  
F;52-117/Domain: carboxyl-terminal propeptide #status predicted <CTP>  
F;26/binding site: octanote (Ser) (covalent) #status experimental

Query Match 32.1%; Score 198; DB 1; Length 117;  
Best Local Similarity 88.6%; Pred. No. 1.6e-13;  
Matches 39; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MPSPGTVCSLLGLMLDLAMAGSSFLSPHQVQVRPPHKAP 44  
DB 1 MPSPGTVCSLLGLMLDLAMAGSSFLSPHQVQVRPPHKAP 44

RESULT 2

B59316  
ghrelin precursor - rat  
N;Alternate names: preproghrelin  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 16-Jun-2000 #sequence\_revision 16-Jun-2000 #text\_change 09-Jul-2004  
C:Accession: B59316  
R;Kojima, M.; Hosoda, H.; Date, Y.; Nakazato, M.; Matsuo, H.; Kangawa, K.  
Nature 402, 656-660, 1999  
A;Title: Ghrelin is a growth-hormone-releasing acylated peptide from stomach.  
A;Reference number: A59316; MUID:20067959; PMID:10604470  
A;Accession: B59316  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA; protein  
A;Residues: 1-117 <KOJ>  
A;Cross-references: UNIPROT:Q9QYH7; GB:AB029433; NID:G6691569; PIDN:BAA89370.1; PID:G66  
A;Experimental source: strain SD; tissue stomach endocrine cells

A;Note: submitted to GenBank, June 1999  
 C;Comment: Ghrelin secreted by the stomach stimulates the release of somatotropin (growth hormone) from the pituitary gland.  
 C;Superfamily: motilin  
 C;Keywords: hormone; lipopeptide; stomach  
 F;1-23/Domain: signal sequence #status predicted <SIG>  
 F;24-51/Product: ghrelin #status predicted <MAT>  
 F;52-117/Domain: carboxyl-terminal propeptide #status predicted <CTP>  
 F;26/Binding site: octanoate (Ser) (covalent) #status experimental

Query Match 25.9%; Score 159.5; DB 1; Length 117;  
 Best Local Similarity 40.0%; Pred. No. 1.7e-09;  
 Matches 42; Conservative 7; Mismatches 33; Indels 23; Gaps 2;

QY 1 MSPSGTVCSSLLGLMLDLAMAGSFLSPHQVQVPPHKAHVVPALPLSNQLCDLE 60  
 DB 1 MYSSATICSLLLSMLMDMAMAGSFLSPHQVQVPPHKAHVVPALPLSNQLCDLE 54

QY 61 QQRH-----WASVFSQSTKDSGDLTVSGRTWG 88  
 DB 55 GWLHPEDRGQAEEAELEIRFNAPFDVGIKLSGAQYQQHGRALG 99

RESULT 3  
 JC4335  
 anti-mullerian hormone type II receptor precursor - human  
 C;Species: Homo sapiens (man)  
 C;Date: 06-Dec-1995 #sequence\_revision 08-Feb-1996 #text\_change 16-Aug-2004  
 C;Accession: JC4335  
 R;Visser, J.A.; McLuskey, A.; van Beers, T.; Weghuis, D.O.; van Kessel, A.G.; Grootegeed, Biochem. Biophys. Res. Commun. 215, 1029-1036, 1995  
 A;Title: Structure and chromosomal localization of the human anti-mullerian hormone type II receptor  
 A;Reference number: JC4335; MUID:96028015; PMID:7488027  
 A;Accession: JC4335  
 A;Molecule type: mRNA  
 A;Residues: 1-573 <VIS>  
 A;Cross-references: UNIPROT:Q16671; GB:X91156; NID:g1107671; PIDN:CAA62593.1; PID:e19804  
 C;Comment: This is a receptor for anti-mullerian hormone (see PIR:WFHUM). It plays a critical role in the development of the male reproductive system.  
 C;Genetics:  
 A;Gene: GDB:AMHR2  
 A;Map position: 12q13-12q13  
 A;Introns: 16/3; 77/3; 141/3; 167/3; 207/2; 284/2; 322/3; 380/2; 429/3; 475/2  
 C;Superfamily: protein kinase homology  
 C;Keywords: ATP; hormone receptor; transmembrane protein  
 F;1-16/Domain: signal sequence #status predicted <SIG>  
 F;17-573/Product: anti-mullerian hormone type II receptor #status predicted <MAT>  
 F;17-141/Domain: extracellular hormone binding #status predicted <ELB>  
 F;142-167/Domain: transmembrane #status predicted <TM>  
 F;201-512/Domain: protein kinase homology <KIN>

Query Match 12.3%; Score 75.5; DB 2; Length 573;  
 Best Local Similarity 28.7%; Pred. No. 5.7;  
 Matches 33; Conservative 15; Mismatches 44; Indels 23; Gaps 3;

QY 3 SPGTVCSS-----LLLLGLMLDLAMAGS---SFLSPHQVQVPPHKAHV 45  
 DB 128 SPGTGSGQQAAGSGSIWALVLLGLFLLLLVLSIILALQKNYVRGEF----- 181

QY 46 VVPALPLSNQLCDLEQQRHWSVFSQSTKDSGDLTVSGRTWGLRVNRLFPSS 100  
 DB 182 VPEPRPSGRDWSVELQELPELCPFSQVIREGGHAVVWAGLOGLKLVAKFPPRS 236

RESULT 4  
 G84707  
 probable MYB family transcription factor [imported] - Arabidopsis thaliana  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
 A;Reference number: A84420; MUID:20083487; PMID:10617197  
 A;Accession: G84707  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-519 <STO>  
 A;Cross-references: UNIPROT:Q8RYE3; GB:AE002093; NID:96598822; PIDN:AAB63073.2; GSPDB:GSPDB:1  
 C;Genetics:  
 A;Gene: At2g30380  
 A;Map position: 2

Query Match 11.8%; Score 72.5; DB 2; Length 519;  
 Best Local Similarity 26.9%; Pred. No. 10;  
 Matches 21; Conservative 13; Mismatches 27; Indels 17; Gaps 5;

QY 17 WLDLWAGSFLSPHQVQVPPHKAHVVPALPLSNQLCDLEQQRHWSV 69  
 DB 431 WYDL-VTGKILALDLHTEQFDVPRPHSGPTPLANLEDRLAMICDYEFSEHVKVAIW 489

QY 70 SOSTKDSGDLTVSGRTW 87  
 DB 490 TMDTQ-----ETW 498

RESULT 5  
 T22943  
 hypothetical protein F58G11.3 - *Caenorhabditis elegans*  
 C;Species: *Caenorhabditis elegans*  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C;Accession: T22943  
 R;Percy, C.  
 submitted to the EMBL Data Library, October 1996  
 A;Reference number: Z19640  
 A;Accession: T22943  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-666 <WIL>  
 A;Cross-references: UNIPROT:P90898; EMBL:Z81094; PIDN:CA803154.1; GSPDB:GN00023; CESP:FP:5  
 C;Experimental source: clone F58G11  
 C;Genetics:  
 A;Gene: CBSP:F58G11.3  
 A;Map position: 5  
 A;Introns: 42/2; 82/2; 153/3; 274/3; 380/1; 569/3; 613/3

Query Match 11.8%; Score 72.5; DB 2; Length 666;  
 Best Local Similarity 26.9%; Pred. No. 14;  
 Matches 32; Conservative 10; Mismatches 42; Indels 35; Gaps 5;

QY 33 QRQVQRPHPKAPHVVPALPLSNQLCDLEQQRHWSVFSQSTKDSGSD----- 79  
 DB 537 QRVRVNPQCVKVKVVPVTLGQALAEVRKQEQVEAFNEQPEPSPRLGMSSSHAA 596

QY 80 LTVSGRTWGLRVNLR-----LFPSS-----RERSR-----RSHQPSCSPEL 116  
 DB 597 SNVSDGGAQVQVQVKEKSPKPPFTVLLPPMKGAGVKIRPRSVLVCHSSASSFPFSL 655

RESULT 6  
 E84211  
 virulence-associated protein vacB homolog - *Mycoplasma genitalium*  
 C;Species: *Mycoplasma genitalium*  
 C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
 C;Accession: E84211  
 R;Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J. Science 270, 397-403, 1995  
 A;Title: The minimal gene complement of *Mycoplasma genitalium*.  
 A;Reference number: A84200; MUID:96026346; PMID:7569993  
 A;Accession: E84211  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-725 <TIGR>

A;Crose-references: UNIPROT:P47350, GB:U39690, GB:L43967; NID:G1045782; PID:G1045783; TID:G1045784  
A;Genetic code: SGC3  
C;Superfamily: virulence-associated protein vacB homolog  
C;Genetics:  
A;Experimental source: strain G-37

Query Match	11.8%;	Score 72.5;	DB 1;	Length 725;
Best Local Similarity	26.7%;	Pred. No. 15;		
Matches 27;	Conservative 16;	Mismatches 37;	Indels 21;	Gaps 4;

[illegible]

```
Qy      68 VFSQSTKDSGSDLTVSGRTWGLRVLNRLFPSSRRSRSSH 108
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      358 VCEISFDNOGRIKT-----NKLYPATIIISKNRFSY 387
```

RESULT 7  
148882  
thyrotropin receptor precursor - mouse  
N:Alternate names: thyroid-stimulating hormone receptor; TSH receptor  
C:Species: Mus musculus (house mouse)  
C:Date: 15-Mar-1996 #sequence\_revision 15-Mar-1996 #text\_change 09-Jul-2004  
C:Accession: I48882  
R:Stein, S.A.; Oates, E.L.; Hall, C.R.; Grumbles, R.M.; Fernandez, L.M.; Taylor, N.A.; F  
Mol. Endocrinol. 8, 129-138, 1994  
A:Title: Identification of a point mutation in the thyrotropin receptor of the hyt/hyt R  
A:Reference number: A54371; MUID:94224232; PMID:8170469

A:Accession: 140002  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-764 <RES>  
A:Cross-references: UNIPROT:P47750; EMBL:U02602; NID:G575923; PID:G575923  
C:Genetics:

A:Gene: TSHR  
C:Superfamily: glycoprotein hormone receptor; leucine-rich alpha-2-glycoprotein repeat  
F:Keywords: G protein-coupled receptor; transmembrane protein  
F:53-76/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR1>  
F:77-101/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR2>  
F:102-126/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR3>  
F:127-151/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR4>  
F:152-176/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR5>  
F:179-200/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR6>  
F:201-226/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LR7>

Query Match 11.7%; Score 72; DB 2; Length 764;  
Best Local Similarity 29.5%; Pred. No.18;  
Matches 33; Conservative 15; Mismatches 48; Indels 16; Gaps 5;

QY 4 PGTCSLLLLGWLWLDLAMAGSSFLSP--HQRVQVRPPHKAPHVVPALPLSNQLCDLE 60

Db 3 PGSL--LLLVLLALSRSLRGKECASPCCECHOEDDFRYTCKEHRIPSIPPSSTQIKL- 59

Qy 61 QQRHWASV-----FSQSTKDSGSDLTVSGRTWGLRVLNRLEFPPSSRRSRSH 108  
: | :: || | | : : || | | : : : :  
Db 60 IETHLKITPSLAFESLPNIGRIYLSIDA-----TLORLEPHSFYNIKSMTH 105

```

RESULT 8
T33503
hypothetical protein C50E10.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T33503
R: Nelson, J.; Klinke, B.; Wohldmann, P.; Bauer, C.
Submitted to the EMBL Data Library, October 1998
A: Description: The sequence of C. elegans cosmid C50E10.

```

A;Accession: 133503  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA

A:Residues: 1-307 <NRL>  
A:Cross-references: UNIPROT:Q965H8; EMBL:AF098988; PIDN:AAC67435.1; GSPDB:GN000020; CESP  
A:Experimental source: strain Bristol N2; clone C50E10  
C:Genetics:  
A:Gene: CESP:C50E10.3  
A:Map position: 2  
A:Introns: 232/3

Query Match 11.4%; Score 70; DB 2; Length 307;  
Best Local Similarity 29.9%; Pred. No. 11;  
Matches 23; Conservative 13; Mismatches 27; Indels 14; Gaps 4;

QY 10 LLLLCWLWLDIAMAGSSFLSEHQVQVRPPHKAHVVPALPSNQLCD-LEQ--QRHWA 66  
| : | | | : | | : : | | : : | | : : | | : : | | : : | | : : |  
Db 209 LIFLGNRWFWNLKIYNVEFLERPE----EAKPSSH-----LPGAHSLCEHLFVPAALHPA 25

Qy 67 SVFSQSTKDSGSDLTVS 83  
| : | : | : | : | : | :  
pb 258 SIAPQANKRPSTATIN 274

## RESULT 9

H96792  
unknown protein Fl4G6.10 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: H96792  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.  
ansen, N.F.; Hughes, B.; Huizarr, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: H96792  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-302 <STO>  
A:Cross-references: UNIPROT:Q9C9K7; GB:AE005173; NID:g6642668; PIDN:AAF20248.1; GSPDB:G.  
C:Genetics:  
A:Gene: Fl4G6.10  
A:Map position: 1  
C:Superfamily: Arabidopsis thaliana hypothetical protein T12H17.200

Query Match 11.3%; Score 69.5; DB 2; Length 302;  
Best Local Similarity 31.9%; Pred. No. 12;  
Matches 30; Conservative 6; Mismatches 31; Indels 27; Gaps 6

QY 30 PE-HQRVQVRPP-HKAPHVVPALPLSNQLCDLEQQRHWASVFSQSTKDSGSDLTVSGRTW 87

pb 21 PEYHLOLOPOLHPYPOPOPOPOPOONSDE-----SDSNKPCGSDPVTSGST- 70

Qy 88 GLRVNRLFPSSRRSR-----RSH 108  
| | | | | : : :  
Db 71 GKRPGR--PPGSKNKPDPVIVTRDSPNVI.RSH 102

DEPOSIT 10

hypothetical 51.8K protein - chicken anemia virus  
C39926  
N:Alternate names: ORF 3  
C:Species: chicken anemia virus, CAV  
C:Date: 17-Jan-1992 #sequence\_revision 17-Jan-1992 #text\_change 09-Jul-2004  
C:Accession: C39926; C48343; JQ1175  
R:Notebook, M.H.M.; de Boer, G.F.; van Roozelaar, D.J.; Karreman, C.; Kranenburg, O.; V.  
J. Virbol, 65, 3111-3139, 1991  
A:Title: Characterization of cloned chicken anemia virus DNA that contains all elements  
A:Reference number: A39926; MUID:91237831; PMID:1851873  
A:Accession: C39926

A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-449 <NOT>  
A:Cross-references: UNIPROT:Q99153; GB:M55918; NID:g323250; PIDN:AAA91824.1; PID:g323253  
R:Meenan, B.M.; Todd, D.; Creelan, J.L.; Earle, J.A.; Hoey, E.M.; McNulty, M.S.  
Arch. Virol. 124, 301-319, 1992  
A>Title: Characterization of viral DNAs from cells infected with chicken anaemia agent:  
A:Reference number: A48343; MUID:92296898; PMID:1605740  
A:Accession: C48343  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-13, 'A', 15-28, 'K', 30-253, 'E', 255-264, 'N', 266-320, 'R', 322-449 <MBE>  
A:Cross-references: GB:M81223; NID:g323254; PIDN:AAA42884.1; PID:g323257  
A:Experimental source: isolate Cux-1  
A>Note: sequence extracted from NCBI backbone (NCBIN:106168, NCBI:P106171)  
R:Claessens, J.A.J.; Schrier, C.C.; Mockett, A.P.A.; Jagt, E.H.J.M.; Sondermeijer, P.J.A.  
J. Gen. Virol. 72, 2003-2006, 1991  
A>Title: Molecular cloning and sequence analysis of the genome of chicken anaemia agent.  
A:Reference number: JQ1175; MUID:91341490; PMID:1908516  
A:Accession: JQ1175  
A:Molecule type: DNA  
A:Residues: 1-13, 'A', 15-143, 'E', 145-156, 'M', 158-250, 'R', 252-286, 'T', 288-445, 'E', 447-449  
A:Cross-references: GB:D10068; GB:D01218; NID:g221116; PIDN:BAA00957.1; PID:g221118  
A:Experimental source: strain 26P4  
A>Note: the authors translated the codon AYT for residue 261 as Ile and ARC for residue  
C:Comment: Chicken anemia agent causes infectious anemia and immunosuppression in young  
C:Comment: This sequence is encoded by the replicative form DNA.  
C:Superfamily: chicken anemia virus 52K hypothetical protein  
Query Match 11.1%; Score 68.5; DB 2; Length 449;  
Best Local Similarity 27.5%; Pred. No. 23;  
Matches 22; Conservative 9; Mismatches 38; Indels 11; Gaps 3;  
QY 28 LSPHQVQVPPHKKAPHVVPALPLSNQLCDLEQQRHWASVFSQSTKDSGLTVSGRTW 87  
DB 268 MRPDEQVAMP--DPPIITATTAGTQVCWVNSTQAW---WSWDVTYMSFATLTALGAQW 323  
QY 88 GLRVNLRLFPSSRSRERS 107  
DB 324 S-----FPFGQRSVRSRS 336  
RESULT 11  
T35474  
50KD proline rich protein - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C>Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 09-Jul-2004  
C:Accession: T35474  
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, December 1998  
A:Reference number: Z21579  
A:Accession: T35474  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-456 <OLI>  
A:Cross-references: UNIPROT:Q9ZBP2; EMBL:AL034492; PIDN:CAA22501.1; GSPDB:GN000070; SCOE  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOE:DB:SC6C5.03  
Query Match 11.1%; Score 68.5; DB 2; Length 456;  
Best Local Similarity 27.9%; Pred. No. 24;  
Matches 34; Conservative 10; Mismatches 63; Indels 15; Gaps 5;  
QY 2 PSFGTVCSLLGLMLWLDLAMA---GSSFLSPHQVQVPPH--KAPHVVPALPLSNQL 56  
DB 39 PARGVARSAAFGPWSGARGARHFGARPTPGDRRRYRPPHGNRLPEPVG-PRHR-- 95  
QY 57 CDLEQQRHWASVFSQSTKDSGLTVSGRTWGLRVNLRLFP-----PSRERSRHSHQSPC 112  
DB 96 ---HOORRGTPAVGTGPEAPASTTTRRAGPAPAVRRAGHRSRPRTGPKRKHKTCSRFQSP 152  
QY 113 SP 114

Db 153 SP 154

## RESULT 12

T32425

hypothetical protein C10E2.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C&gt;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T32425

R:Wohlmann, P.; Sansone, J.

submitted to the EMBL Data Library, September 1997

A:Description: The sequence of C. elegans cosmid C10E2.

A:Reference number: Z21165

A:Accession: T32425

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-796 &lt;WOH&gt;

A:Cross-references: UNIPROT:O17323; EMBL:AF026202; PIDN:AAB71243.1; GSPDB:GN000028; CESP:

A:Experimental source: strain Bristol N2, clone C10E2

C:Genetics:

A:Gene: CESP:C10E2.3

A:Map position: X

A:Introns: 85/2; 220/2; 269/1; 305/1; 519/3; 576/3; 724/3; 755/3

Query Match 11.1%; Score 68.5; DB 2; Length 796;

Best Local Similarity 28.9%; Pred. No. 44;

Matches 28; Conservative 15; Mismatches 35; Indels 19; Gaps 4;

QY 18 LDLMAGSGSFLSPHQVQV-RPPHKKAPHVVPALPLSNQLCDLEQQRHWASVFSQSTKDS 76

DB 47 LSLANSLTNLLSSNGNLSPQPTTKEHH--PTAPTSNRKCDLPSRN--STTISQLTKD- 101

QY 77 GSDLTVSGRTWGLRVNLRLFPSSRERSRHSHQSPSCS 113

DB 102 -----RLKNMIANRSKGSNSQSNLMSNS 125

## RESULT 13

B75564

hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C&gt;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C:Accession: B75564

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

Science 286, 1571-1577, 1999

A&gt;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: B75564

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-94 &lt;WHI&gt;

A:Cross-references: UNIPROT:Q9RY91; GB:AE001869; GB:AE000513; NID:g6457721; PIDN:AAF0961

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0059

A:Map position: 1

Query Match

Best Local Similarity 23.9%; Pred. No. 4.7;

Matches 22; Conservative 12; Mismatches 30; Indels 28; Gaps 3;

QY 32 HQRVQVPPHKKAPHVVPALPLS-----NQLCDLEQQRHWASVFSQSTKDSGL 80

DB 5 HQHQQAQRGHRQHRDATHPLAQPEQQQRHHEQQQQQAHSARLQQRKSAG--- 61

QY 81 TVSGRTWGLRVNLRLFPSSRERSRHSHQSPSC 112

DB 62 --TGQWNG-----RERNGISQCTC 79



THIS PAGE LEFT BLANK